

PROYECTO INTEGRADOR

CASO DE ESTUDIO: PRESENTISMO DE TURNOS EN INSTITUCIÓN MÉDICA

GRUPO 9

INTEGRANTES:

- IGNACIO MACHADO
- ANDREA LORENA VILLALBA
- LAUTARO PEREIRA BASILE
- FEDERICO GURNÉ

PRESENTACIÓN DEL CASO: PRESENTISMO EN INSTITUCIÓN MÉDICA

- Se cuenta con base de datos de turnos de pacientes de una hospital indicando si asistió al turno o no.
- Se nos solicita realizar un clasificador que pueda predecir cuando un paciente faltará al turno.
- Es un punto muy crítico de los hospitales ya que hay un costo de oportunidad muy grande si el paciente falta a su turno.

Plan de Trabajo

1. Realizar análisis exploratorio de los datos para lograr un mejor entendimiento del caso.
2. Limpieza de los datos.
3. Establecer un baseline y métricas sobre el cual va a trabajar el estimador.
4. Generar un Clasificador que prediga el target 'No-Show'.
5. Conclusiones.

Descripción del DATASET

- 110.189 turnos
- 12 features (11 + 1)
- Muestra de 3 meses (Abril, Mayo y Junio de 2016)

DESCRICION DEL DATASET

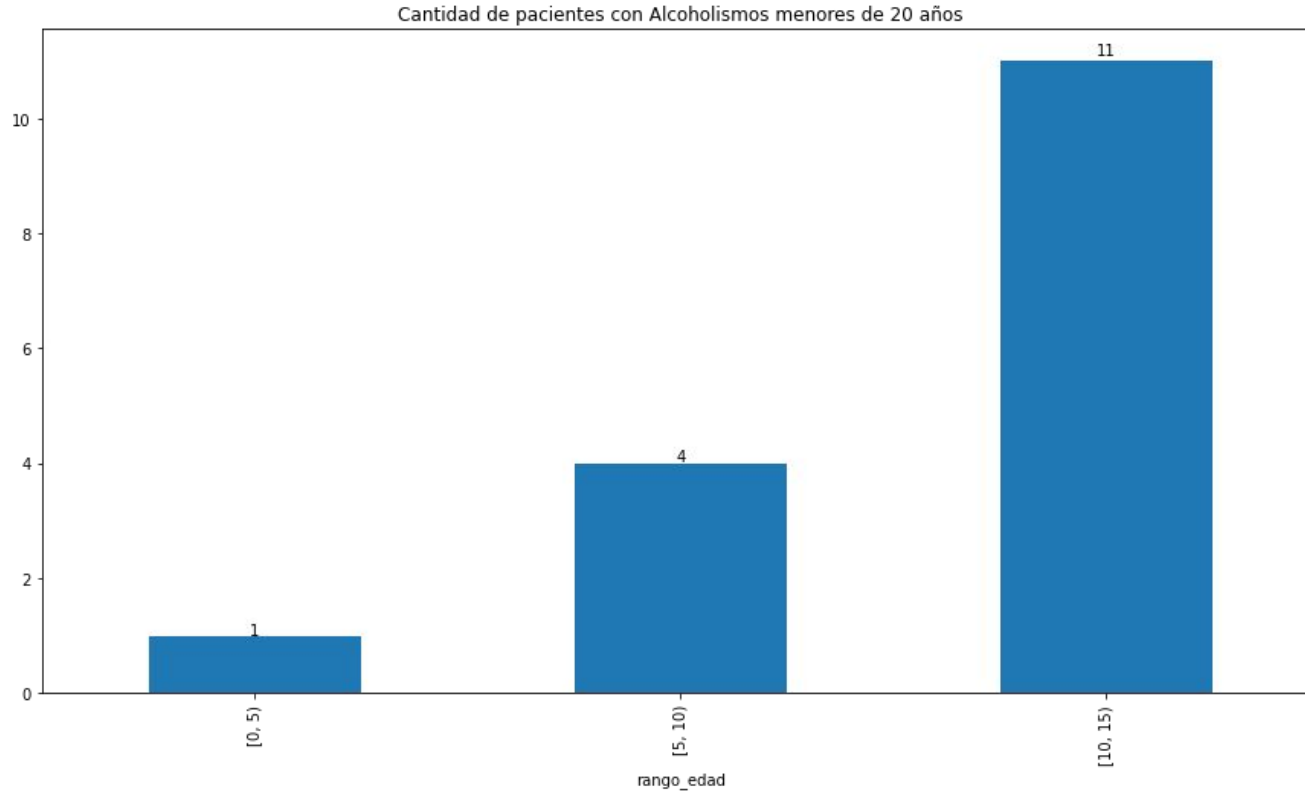
El Dataset a utilizar fue extraido de kaggle referido a una base de datos de un hospital en Brasil

Descripción de Columnas

- . Gender : Sexo -> Variable Explicativa
- . ScheduledDay : Dia en que se agenda el turno -> Variable Explicativa
- . AppointmentDay : Dia del turno -> Variable Explicativa
- . Age : Edad -> Variable Explicativa
- . Neighbourhood :Barrio -> Variable Explicativa
- . Scholarship : Si tiene bolsa familia (Plan Social) -> Variable Explicativa
- . Hipertension : Si el paciente tiene hipertencion -> Variable Explicativa
- . Diabetes : Si el paciente tiene diabetes -> Variable Explicativa
- . Alcoholism : Si el paciente tiene alcoholismo -> Variable Explicativa
- . Handcap : Cantidad de discapacidades que tiene el paciente -> Variable Explicativa
- . SMS_received : Si se le ha enviado un sms de recordatorio -> Variable Explicativa
- . No-show : Si el paciente ha faltado -> Variable Target

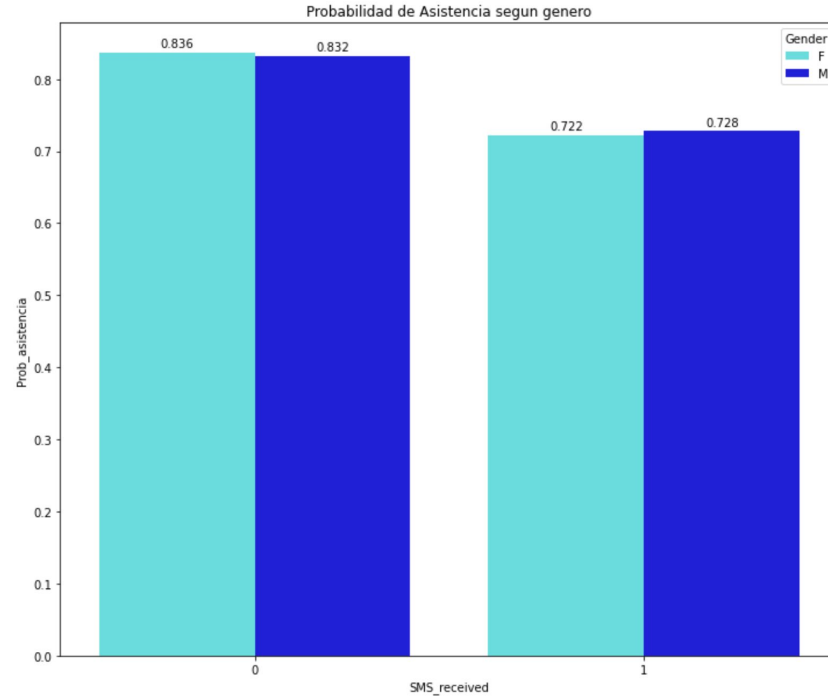
INCONSISTENCIAS EN EL DATASET

Pacientes con Alcoholismo menores a 15 años



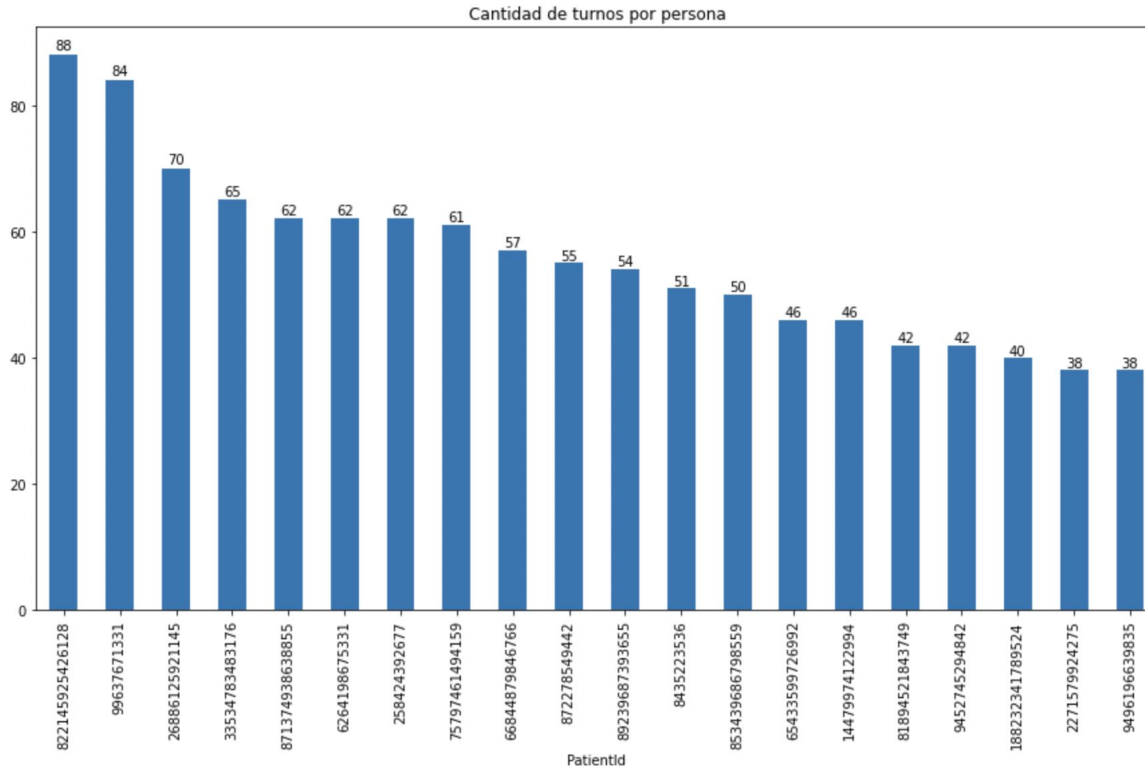
INCONSISTENCIAS EN EL DATASET

INTERCAMBIO DE VALORES EN CAMPO MENSAJE RECIBIDO

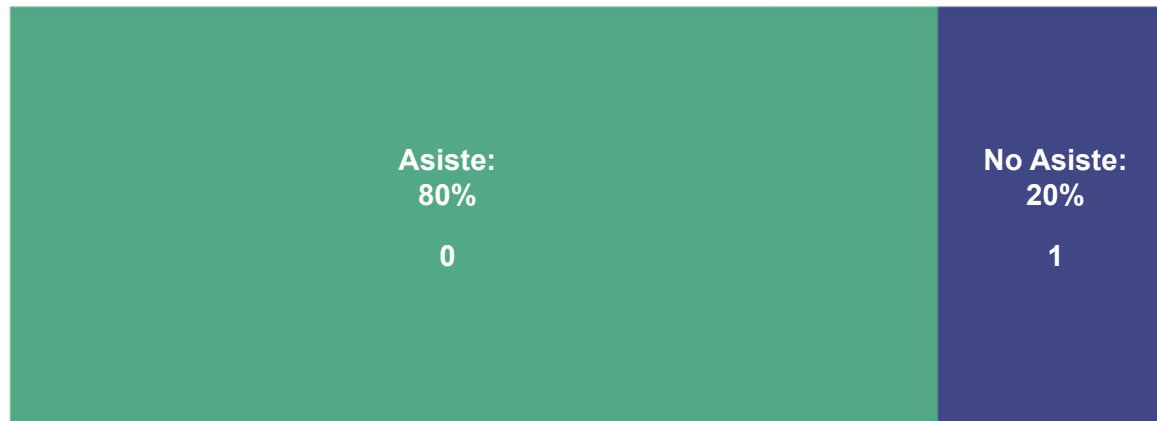


INCONSISTENCIAS EN EL DATASET

Cantidad de turnos por persona en 3 meses



BALANCE DEL TARGET



ANALISIS EXPLORATORIO DE DATOS

CANTIDAD DE TURNOS

110,189

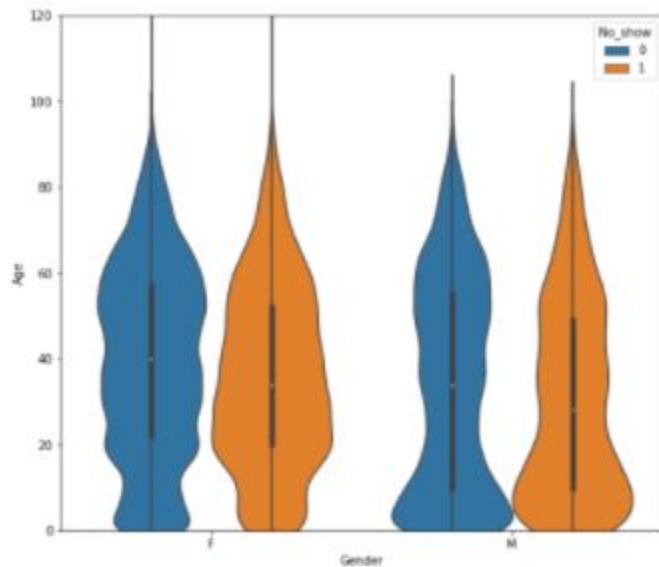
DIAS ENTRE TURNO Y
RESERVA PRESENTE

9

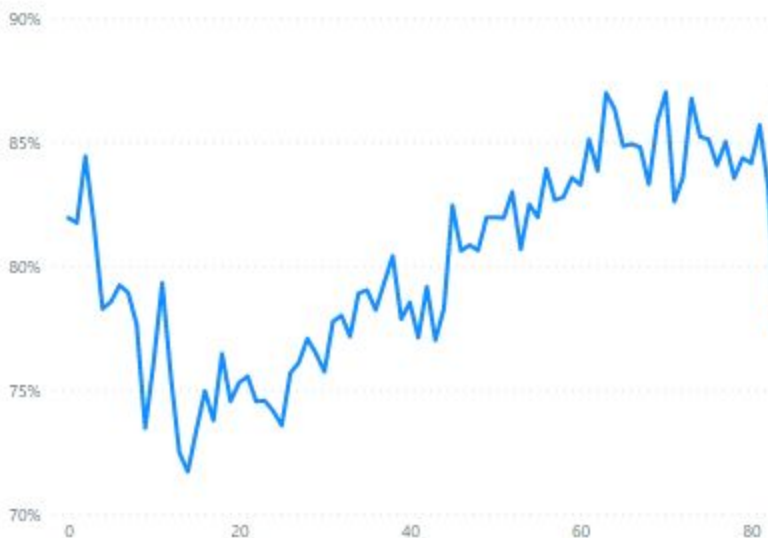
DIAS ENTRE TURNO Y
RESERVA AUSENTES

16

ANALISIS POR EDAD Y PRESENTISMO



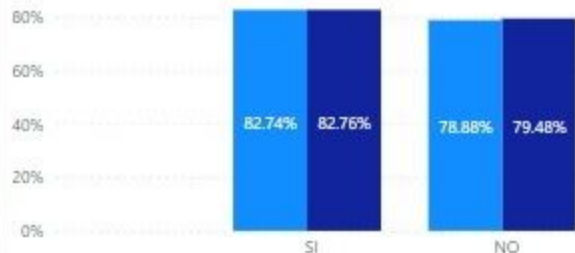
% PRESENTISMO POR EDAD



ANALISIS EXPLORATORIO DE DATOS

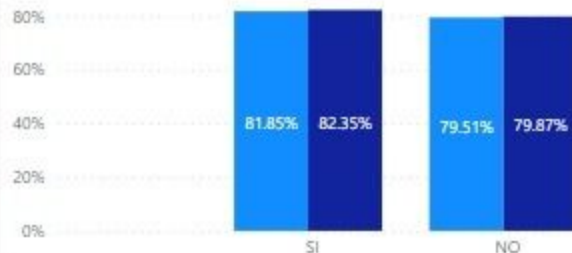
PRESENTISMO POR GENERO E HIPERTENSION

Gender ● F ● M



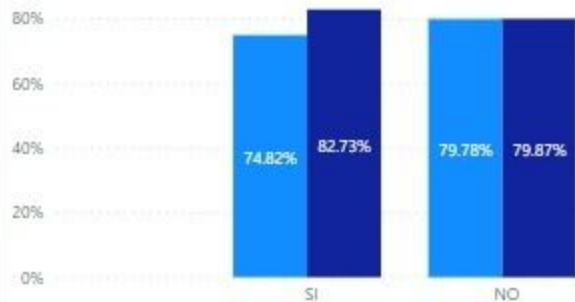
PRESENTISMO POR GENERO Y DIABETES

Gender ● F ● M



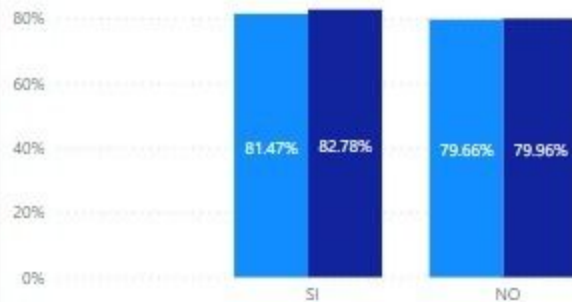
PRESENTISMO POR GENERO Y ALCOHOLISMO

Gender ● F ● M



PRESENTISMO POR GENERO Y DISCAPACIDAD

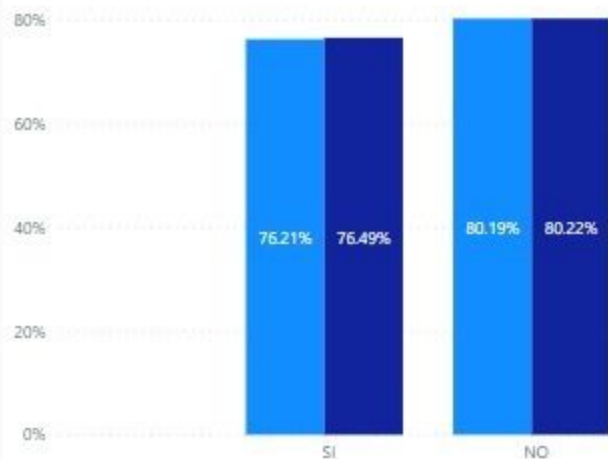
Gender ● F ● M



ANALISIS EXPLORATORIO DE DATOS

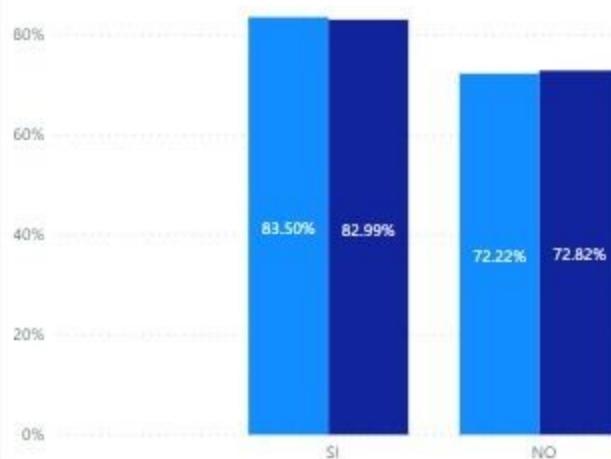
PRESENTISMO POR GENERO Y PLAN SOCIAL

Gender ● F ● M



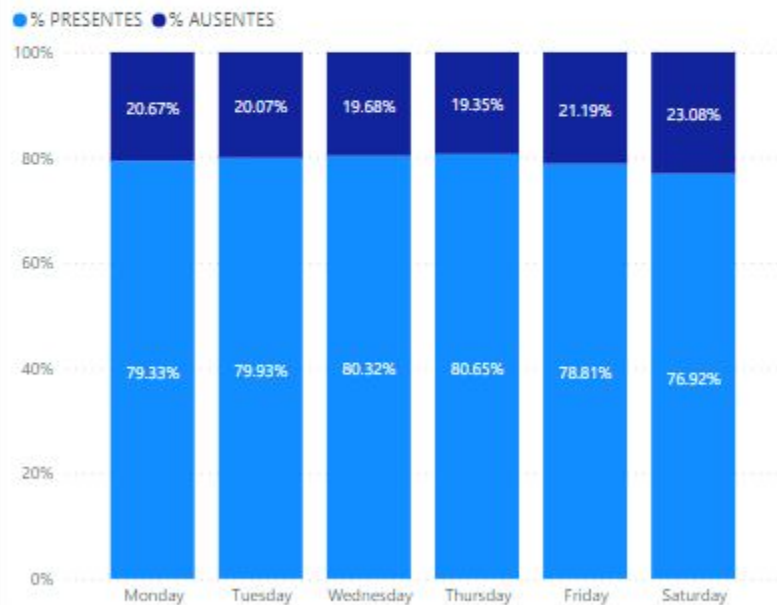
PRESENTISMO POR GENERO Y MENSAJE RECIBIDO

Gender ● F ● M

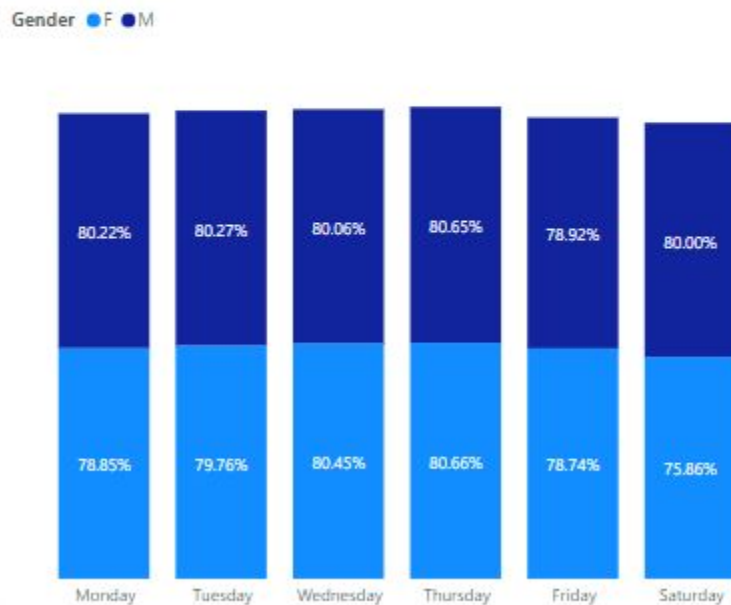


ANALISIS EXPLORATORIO DE DATOS

PRESENTISMO POR DIA DE SEMANA

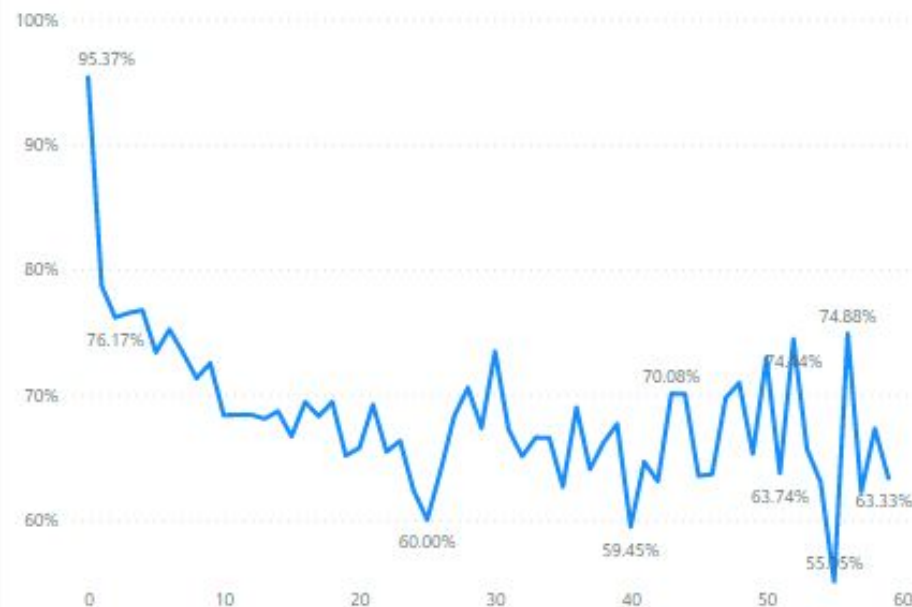


PRESENTISMO POR DIA DE SEMANA Y GENERO

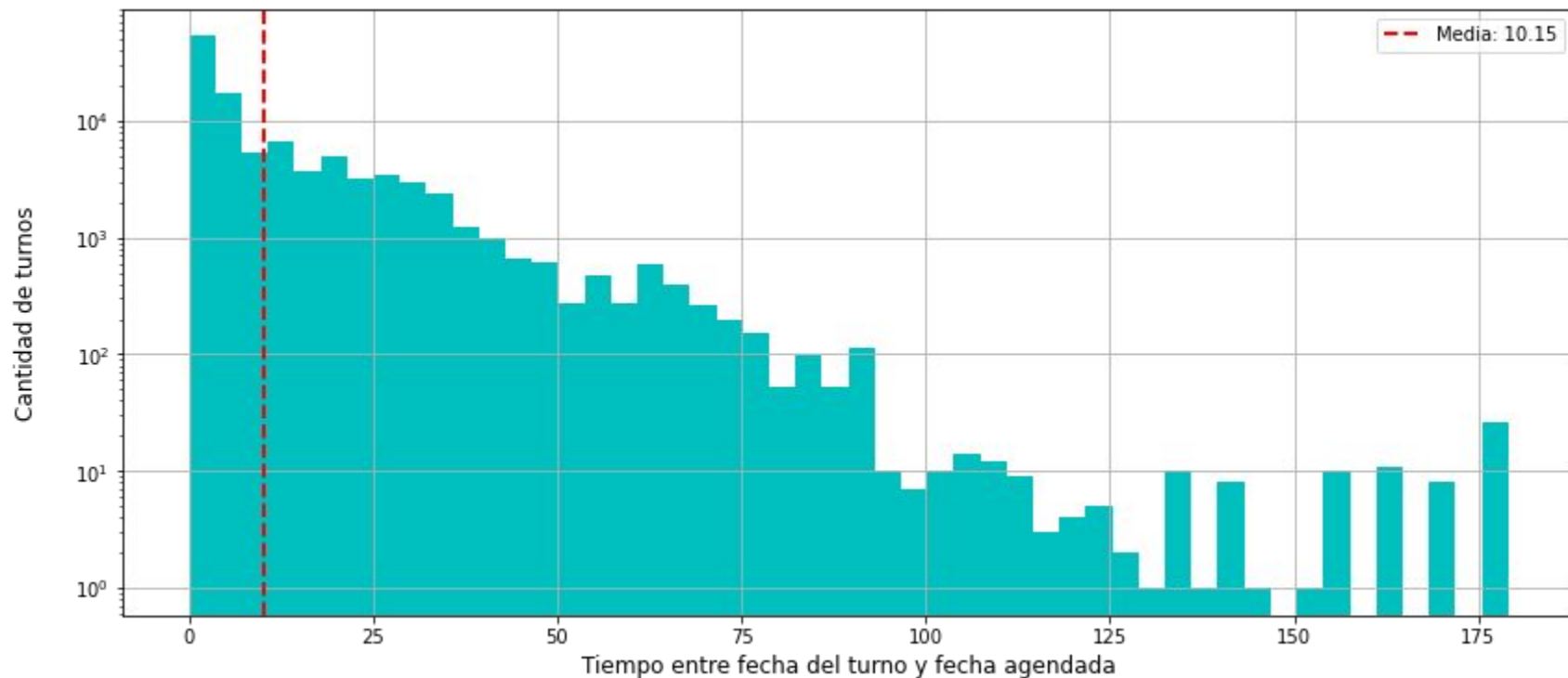


ANALISIS EXPLORATORIO DE DATOS

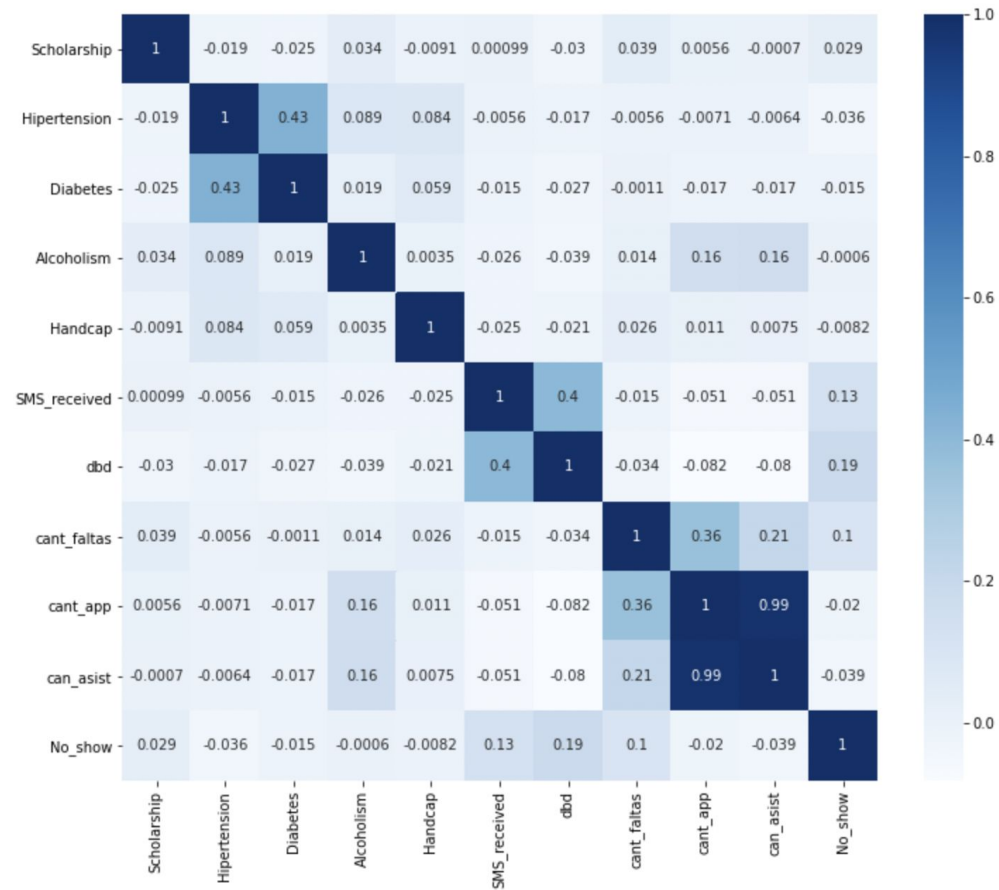
% ASISTENCIA SEGUN DIFERENCIA DIAS TURNO Y RESERVA



Distribución del tiempo entre que se agenda y la fecha del turno

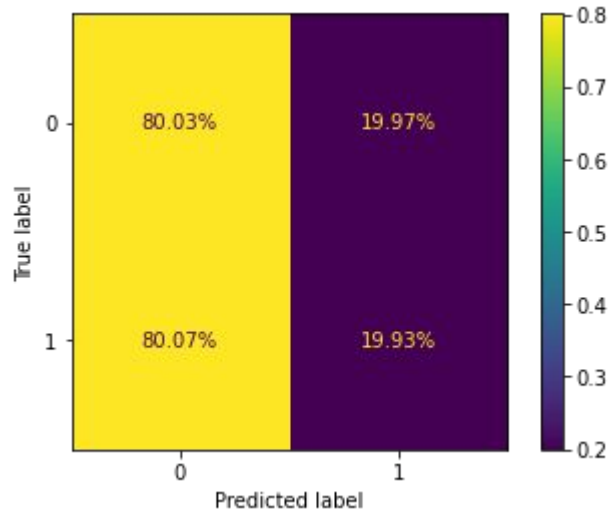


ANÁLISIS DE CORRELACIONES



ANÁLISIS DE MODELOS

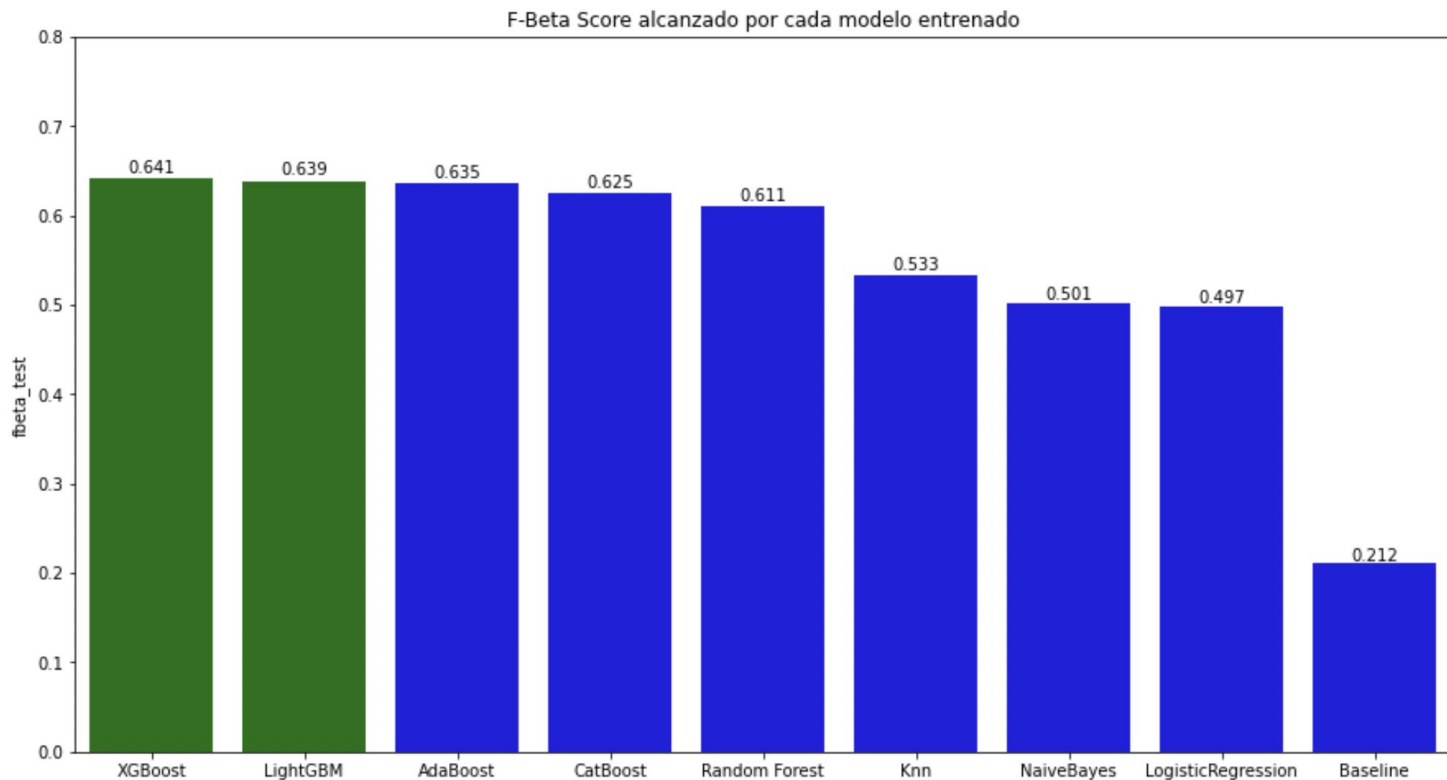
MODELO BASELINE



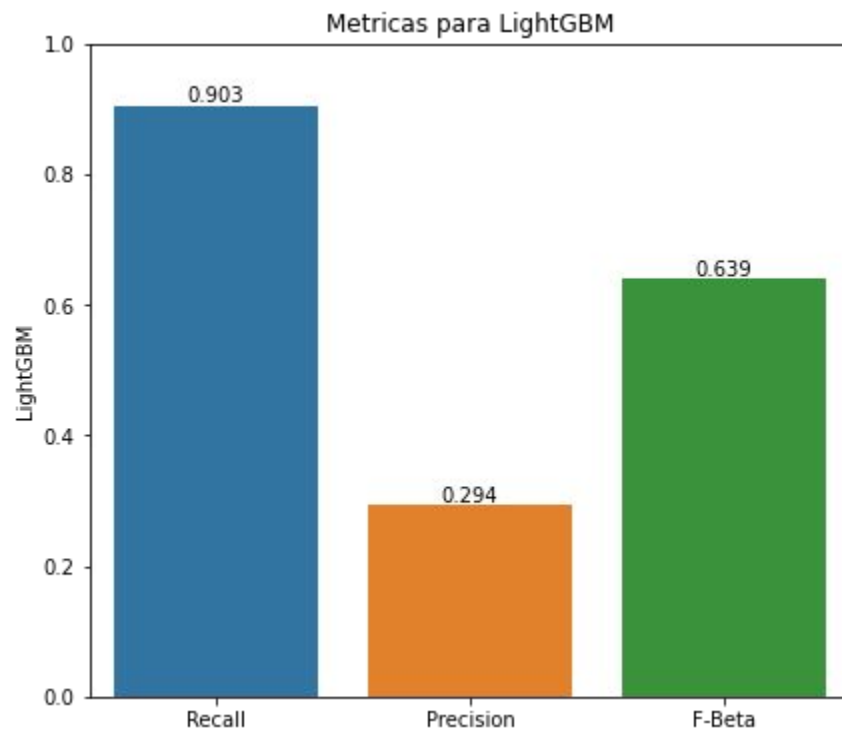
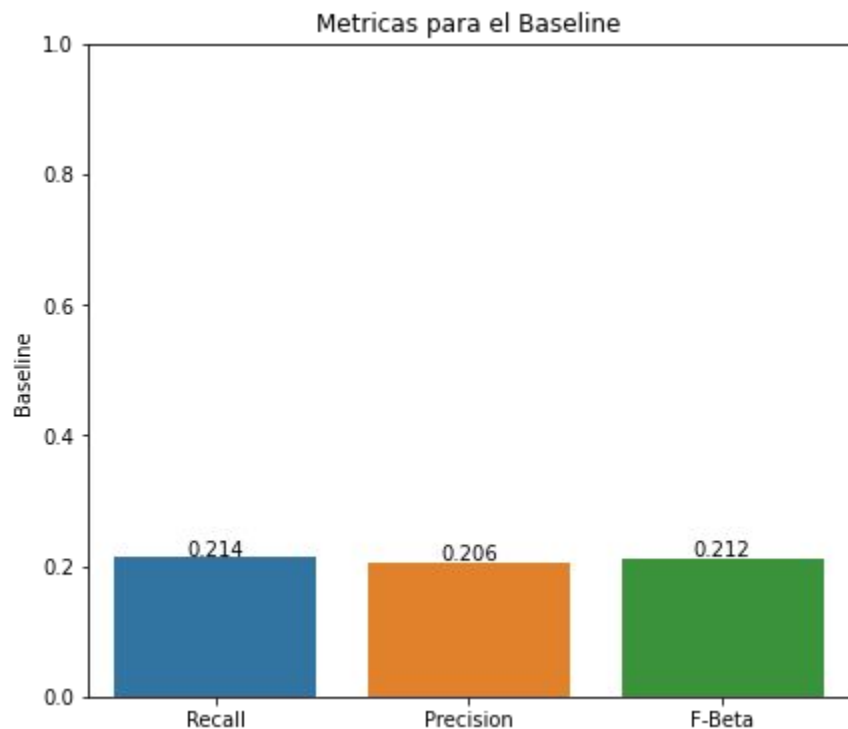
	precision	recall	f1-score	support
0	0.80	0.80	0.80	21890
1	0.20	0.20	0.20	5510
accuracy			0.68	27400
macro avg	0.50	0.50	0.50	27400
weighted avg	0.68	0.68	0.68	27400

Métrica Elegida: F-BETA

Análisis de los distintos modelos entrenados

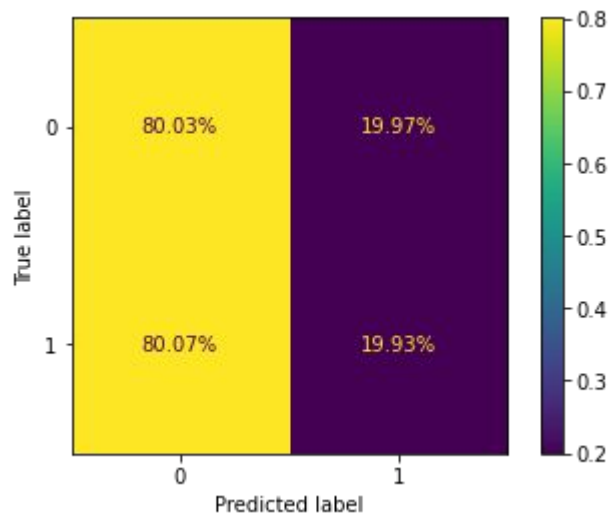


RESULTADO DEL MEJOR MODELO

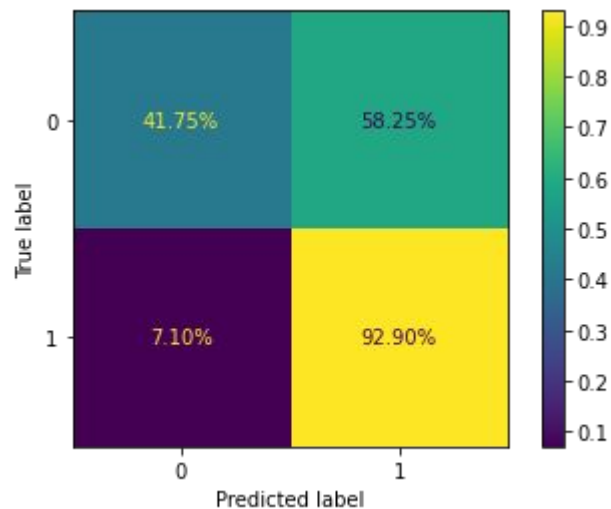


Comparación Baseline - LightGBM

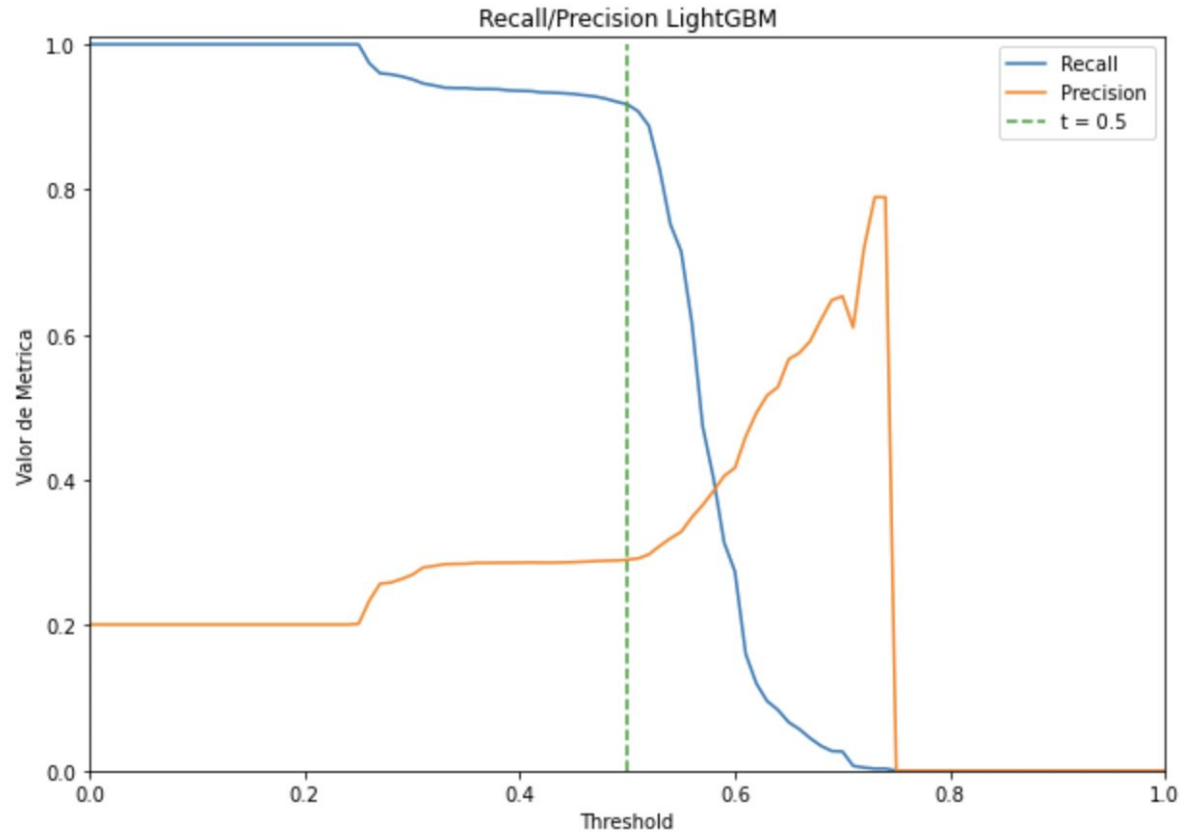
Baseline



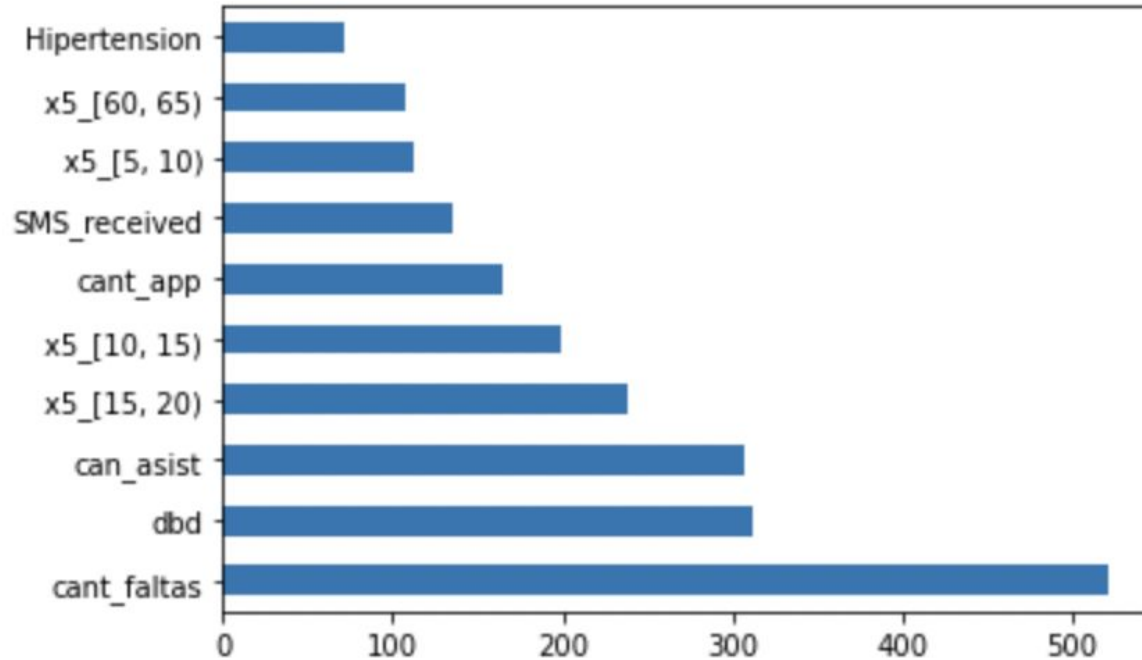
LightGBM



Trade-Off Recall Precision para el Clasificador

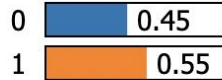


Features más importantes según el modelo

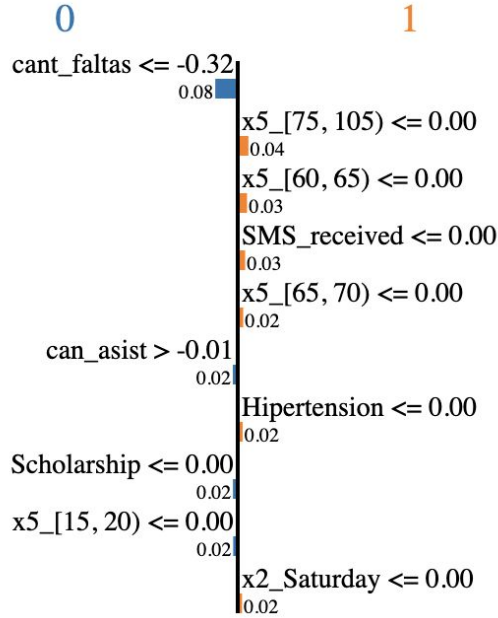


Interpretación del Modelo con Lime

Prediction probabilities

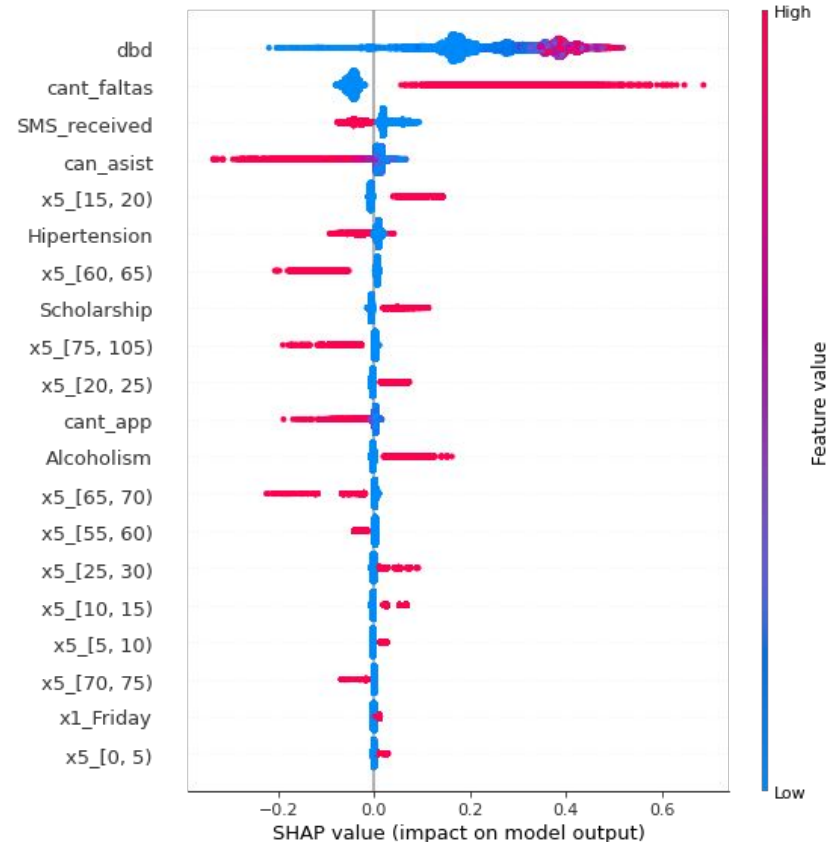
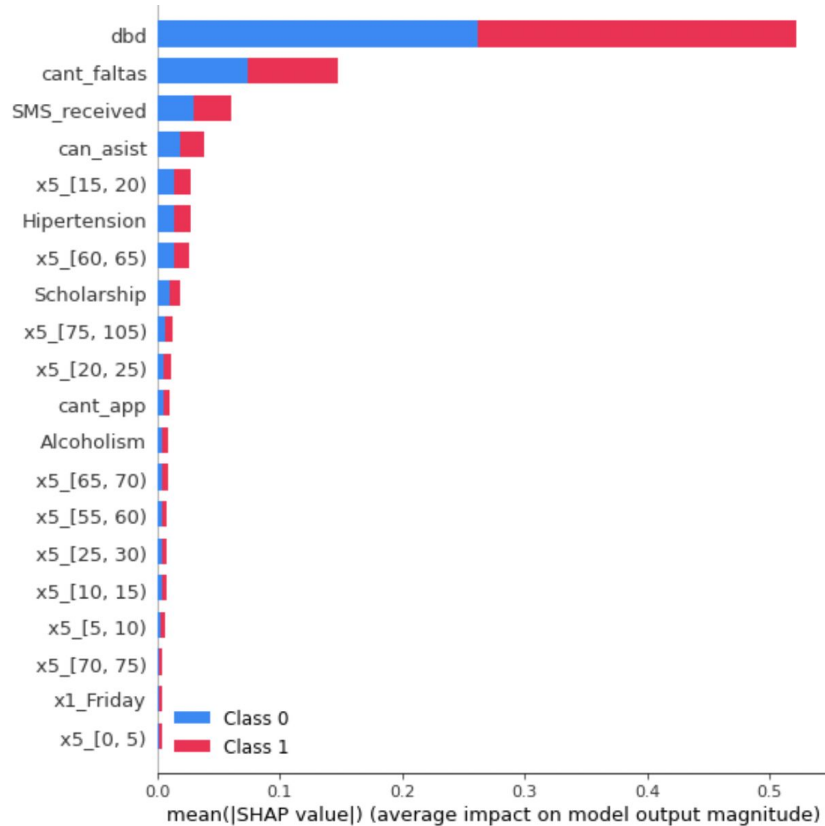


True Label: 0



Feature	Value
cant_faltas	-0.32
x5_[75, 105)	0.00
x5_[60, 65)	0.00
SMS_received	0.00
x5_[65, 70)	0.00
can_asist	0.26
Hipertension	0.00
Scholarship	0.00
x5_[15, 20)	0.00
x2_Saturday	0.00

Interpretación del Modelo con SHAP



Conclusiones

- Las features más importantes para la predicción de si una persona va a faltar son:
 - Tiempo entre fecha del turno y fecha en que se agenda.
 - Si la persona es notificada a modo de recordatorio (en este caso vía SMS).
 - El historial del presentismo del paciente.
 - Si posee hipertensión o no.
- Tenemos una muestra de 3 meses por que lo sería insuficiente para realizar un análisis de series de tiempo y ver la tendencia del presentismo a lo largo del año. (estacionalidad anual)
- Se podrían realizar futuros análisis geoespaciales teniendo la ubicación del centro médico y las ubicaciones de cada paciente para mejorar la precisión de la estimación.
- Obtener la información acerca de las prácticas y servicios referidos al turnos también podría resultar un modelo de mejor performance.
- Segmentar el modelo o estimaciones por grupo etario(?.)